Auto-Gait: Automatic Ataxia Risk Assessment with Computer Vision from Gait Task Videos

WASIFUR RAHMAN* and MASUM HASAN*, University of Rochester, USA

MD SAIFUL ISLAM, University of Rochester, USA

TITILAYO OLUBAJO, Houston Methodist, USA

JEET THAKER, University of Rochester, USA

ABDELRAHMAN ABDELKADER, University of Rochester, USA

PHILLIP YANG, Center for Health + Technology, University of Rochester Medical Center, USA

HENRY PAULSON, University of Michigan, USA

GULIN OZ, University of Minnesota, USA

ALEXANDRA DURR, Hôpital de la Pitié-Salpêtrière, France

THOMAS KLOCKGETHER, University Hospital Bonn, Germany

TETSUO ASHIZAWA AND READISCA INVESTIGATORS, Houston Methodist, USA

EHSAN HOQUE, University of Rochester, USA

Many patients with neurological disorders, such as Ataxia, do not have easy access to neurologists, –especially those living in remote localities and developing/underdeveloped countries. Ataxia is a degenerative disease of the nervous system that surfaces as difficulty with motor control, such as walking imbalance. Previous studies have attempted automatic diagnosis of Ataxia with the help of wearable biomarkers, Kinect, and other sensors. These sensors, while accurate, do not scale efficiently well to naturalistic deployment settings. In this study, we propose a method for identifying ataxic symptoms by analyzing videos of participants walking down a hallway, captured with a standard monocular camera. In a collaboration with 11 medical sites located in 8 different states across the United States, we collected a dataset of 155 videos along with their severity rating from 89 participants (24 controls and 65 diagnosed with or are pre-manifest spinocerebellar ataxias). The participants performed the gait task of the Scale for the Assessment and Rating of Ataxia (SARA). We develop a computer vision pipeline to detect, track, and separate the participants from their surroundings and construct several features from their body pose coordinates to capture gait characteristics such as step width, step length, swing, stability, speed, etc. Our system is able to identify and track a patient in complex scenarios. For example, if there are multiple people present in the video or an

interruption from a passerby. Our Ataxia risk-prediction model achieves 83.06% accuracy and an 80.23% F1 score. Similarly, our Ataxia severity-assessment model achieves a mean absolute error (MAE) score of 0.6225 and a Pearson's correlation coefficient score of 0.7268. Our model competitively performed when evaluated on data from medical sites not used during training. Through feature importance analysis, we found that our models associate wider steps, decreased walking speed, and increased instability with greater Ataxia severity, which is consistent with previously established clinical knowledge. Furthermore, we are releasing the models and the body-pose coordinate dataset to the research community – the largest dataset on ataxic gait (to our knowledge). Our models could contribute to improving health access by enabling remote Ataxia assessment in non-clinical settings without requiring any sensors or special cameras. Our dataset will help the computer science community to analyze different characteristics of Ataxia and to develop better algorithms for diagnosing other movement disorders.

CCS Concepts: • **Applied computing → Health care information systems**; • **Computing methodologies → Object detection**; **Tracking**; **Interest point and salient region detections**.

Additional Key Words and Phrases: ataxia, gait, computer vision, pose estimation, datasets

1 INTRODUCTION

Ataxia is a general term referring to a lack of muscle coordination and control, typically arising due to damage to the cerebellum in the brain [35]. Common symptoms of Ataxia include difficulties in walking, deterioration of fine motor skills, imbalance, and trouble eating and swallowing [47]. In this paper, we analyze two forms of hereditary Ataxia named spinocerebellar ataxia (SCA) types 1 and 3 (SCA1 and SCA3) which demonstrate the fastest progression rate and highest prevalence respectively [22].

Early detection and consistent monitoring of Ataxia can greatly improve symptom management and long-term patient prognosis [35]. The Scale for the Assessment and Rating of Ataxia (SARA) is a popular scale for measuring ataxia progression with high inter-rater reliability, inter-rater consistency, and test-retest reliability [9, 19, 41, 45, 46]. The SARA scale is composed of eight different tasks: gait, stance, sitting, speech disturbance, finger chase, finger-to-nose, fast alternative hand movement, heel-shin slide. These tasks are performed by a participant during a clinical visit in front of a trained neurologist, who observes the patient and assigns them a score for each task. Although this oversight of the evaluation ensures quality, it limits the accessibility and equity of the care. People without access to high-quality healthcare are deprived of this in-person care, and even people with access to this level of care are limited to the frequency of their visits because of the time commitment of each visit. Previous literature on automatic gait analysis for ataxia mostly focused on external sensors, such as wearables, Kinect, etc. [15, 20, 31, 42], which still limits its use-case to the access and setup of the sensors. In this paper, we try to alleviate this problem by developing a computer-vision-based video analysis technique for automatically detecting SCA.

In this study, we have analyzed the gait task in the SARA evaluation metric as, 1) SCA heavily impacts common gait characteristics, and 2) gait provides objective, clinically applicable measurements [8]. We have collected 183 recorded videos of SCA patients performing gait tasks from 11 different medical sites in the US. The videos were annotated with SARA scores by trained neurologists who observed the participant during the walk. Thus, we address the following research questions:

(1) RQ1: Can we provide an early-stage screening of ataxia from the SARA gait task with computer vision?
(2) RQ2: Can we measure the early-stage severity of ataxia from the SARA gait task with computer vision?
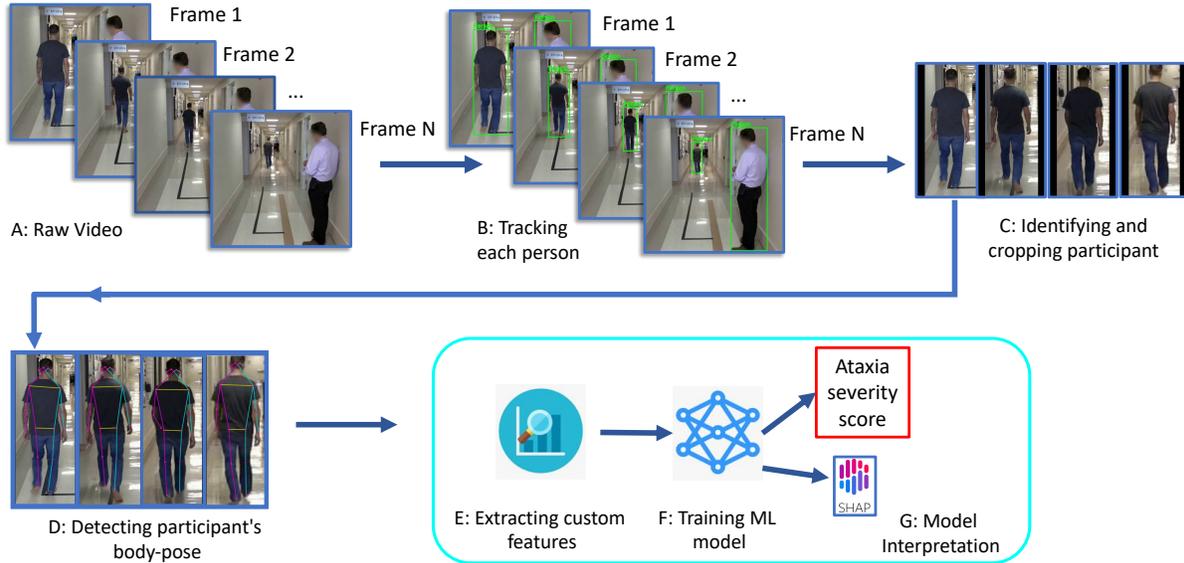
Fig. 1. Ataxia risk/severity prediction pipeline from gait task.

In order to answer those questions, we apply computer vision to quantify different subtle features of the participant's walking. Figure 1 demonstrates our methodology step by step. First, using an object detection and tracking algorithm, we detect every person from a video frame and track them across the video (Fig. 1 B). We develop a robust Subject Separation Algorithm based on the characteristics of the gait task, that separates the participant from several other people in the frame (Fig. 1 C). Furthermore, we measure their movement using a pose estimation model and use this information to build meaningful signals for detecting SCA (Fig. 1 D). For our Risk-prediction task (RQ1), we divided our data into two classes: one with gait task SARA scores of 0 (Not in risk: Class 0) and the other with gait task SARA scores > 0 (In risk: Class 1). We train a machine learning classifier that can differentiate between these two classes with an 83.06% binary accuracy from 10-fold cross-validation (Table 3). For early-stage Severity-assessment (RQ2), we divided our dataset into four classes with gait task SARA scores: 0, 1, 2, > 2. For all data with gait task SARA scores > 2, we replace their label with 3 since we have limited data with gait task SARA scores > 3. Our *Severity-assessment* regression model has a mean absolute error of 0.6225 (STD 0.0132) and a Pearson's correlation coefficient of 0.7268 (STD 0.0144).

For both of these models, we ran a SHAP analysis [33]. Figure 8 *(a)* and Figure 8 *(b)*, corresponding to the *Risk-prediction* and *Severity-assessment* models respectively show that SHAP can provide intuitive, previously-validated results (Sec 5.3). Furthermore, we ran a "leave-one-site-out" analysis, where we exclude data from one site in the training set and test the model's performance on that medical site's data. This allows us to assess how well the model performs on completely unseen data. While the performance of both models decreases through this method, they still perform significantly better than random chance (Table 3) and zero-correlation (Table 5).

Our contributions are as follows:

(1) We collected a large high-quality annotated dataset of patients with Ataxia and a control group. The dataset contains 155 annotated videos from 89 participants that span several years representing different geographic locations and demographics (Table 2). Upon acceptance, we will publicly release an anonymized version of our dataset containing body pose coordinates to improve the existing benchmark for Ataxia

assessment. To the best of our knowledge, it will be the largest accessible dataset on Ataxia analysis from the gait task.

(2) We designed a robust and automated computer vision pipeline that can detect, track, and separate participants regardless of their height, age, gender, and weight. Our pipeline can handle videos across different site settings that include both doctors and other passersby and track their body pose movement.

(3) We developed machine learning algorithms for *Risk-prediction* and *Severity-assessment* of Ataxia from recorded videos of the SARA gait task. Our models achieve 83.06% accuracy and an 80.23 F-1 score for the *Risk-prediction* task (Table 3). For the *Severity-assessment* task, we achieve a mean absolute error (MAE) of 0.6225 and a Pearson's correlation coefficient of 0.7268 (Table 5).

(4) We test our model with completely unseen data using the leave-one-site-out method. Our model generalizes beyond the medical sites they were trained on various lighting conditions, camera resolution, frame rate, and camera positions. (Table 3 and Table 5).

2 RELATED WORKS

Objectively assessing Ataxia features is an active area of interest for many researchers. Although prior efforts mostly focused on using sensors to assess several aspects of Ataxia [28, 37, 55], there is a growing interest in the use of telemedicine-type platforms to increase the accessibility of care for Ataxia patients in non-clinical settings.

2.1 Automated Gait Analysis

Automatically analyzing walking patterns and gait have been studied using various ways. For example, RFID tags [24], wifi signals [58, 60], acoustic signals [59], cellphone [25], accelerometers and sensors[1, 2, 18, 32, 36, 38], Kinect [61], cell-phone [25], etc., all have been shown to be successful in analyzing gait. However, these sensor-based technologies, although accurate in lab settings, sacrifice the accessibility of healthcare. Only a handful of people who own and have the ability to operate these sensors can benefit from such technologies. A more ubiquitous way of sensing gait would be analyzing videos of a participant walking. Previous studies [27, 39] have attempted analysis of gait for clinical application using computer vision. Yet, video-based gait analysis has mostly been explored in a simplistic studio setup – the data is collected from a single site without background noise. Furthermore, the videos are captured from a lateral view requiring a large space, which is unsuitable for homes and even many clinics. In this paper, we introduce a robust video-based gait analysis method trained on videos captured in a more realistic setting. Our method automatically identifies a patient when multiple persons are visible in a video and works across multiple sites. Such generalizability was missing in prior literature.

2.2 Teleneurology for Neurological Disorders

The elderly population is the riskiest group for the most common neurological disorders like Ataxia and Parkinson's disease. Arranging clinical visits for them can be difficult since a clinical visit requires a long commute unsuitable for the elderly, which has been exacerbated by the COVID-19 pandemic. Amid the ongoing public health crisis, experts have recommended greater use of telemedicine to care for patients with such neurological disorders [34]. A shift to telemedicine would be beneficial on two fronts; it would improve the frequency of assessment (therefore painting a more realistic picture of a patient's disease experience), and increase equity by decentralizing care from the clinic. The application of telemedicine in the domain of Parkinson's disease (PD) sets an excellent precedent for this domain [14], especially considering the similarities between PD and Ataxia symptoms. CloudUPDRS [49] is a smartphone application recognized as a Class I medical device by MHRA, UK that captures data from people with PD for future clinical diagnosis. The app employs a deep learning model to ensure the quality of data collected without human supervision [50]. Another smartphone application named mPower [6] can objectively measure PD tremor severity using remotely collected activity data with built-in

smartphone sensors. PARK [30] is a web-based framework that collects webcam videos from subjects for detecting Parkinson's disease. However, compared to PD, remote assessment of Ataxia has received less attention until recently. Lately, Researchers are developing telemedicine care for patients with Cerebellar Ataxia (CA) [34]. Notably, Grobe et al. [17] developed SARA^home, a video-based instrument to measure Ataxia severity from a subject's home. Independently, a user performs 5 of the 8 SARA items; items were modified so that the tool could be easily self-applied without the presence of a provider. After all the tasks are completed, the recorded videos are securely transferred to be evaluated by an Ataxia expert offline using SARA criteria. Separately, Summa et al. [52] created a pilot study for SaraHome, a tool that interlinks the Microsoft Kinect 2.0 and Leap Motion Controller (LMC) for automatic detection of SARA scores. This platform was intended for use by children with early onset Ataxias (EOAs); with help from a caregiver, users complete six of the eight SARA tasks to receive an objective, automated assessment of Ataxia severity [52]. The tool was well-received; however, the aforementioned pilot study was only intended to gauge participant reception and the feasibility of the tool. To date, it has not been tested on actual video footage to assess its performance compared to traditional measures. The emergence of these platforms shows a growing need for automatic home diagnosis tools for CA where our automatic video-based analysis tool would be valuable.

2.3 Detecting Ataxia from Gait Task

The gait task is a significant indicator of Ataxia since patients with CA may demonstrate clumsy, staggering movements while performing the gait task [48]. In the literature, detecting CA from gait has mainly focused on sensor-based analysis. Both wearable sensors and the Kinect system have shown promise in analyzing CA from gait task [15, 20, 26, 31, 42, 53]. LeMoyne et al. [31] gathered accelerometry and gyroscopic data by mounting wireless inertial sensors on the ankle joints of participants and trained a machine-learning model to distinguish subjects with Friedreich's Ataxia from healthy controls. Phan et al. [42] used data from a single gait task to capture features at self-selected slow, preferred, and fast walking speeds for both detecting CA and evaluating its severity. Honda et al. [20] analyzed the gait task by extracting features using a Kinect-based system. Most recently, Dostál et al. [15] obtained accelerometry data using 31 time-synchronized sensors placed on different parts of the body. They obtained an accuracy of 98.0%, and 98.5% in classifying between healthy subjects and subjects with CA, using neural networks extracted features from the shoulders and head/spine respectively. Recently a similar approach obtained 95.8% accuracy in classifying Ataxic and normal gait by using deep learning to analyze frequency components of accelerometric signals simultaneously recorded at multiple body positions. Similarly, Prochazka et al. [43] applied accelerometric signals to train deep learning-based models to achieve 95.8% accuracy in differentiating between ataxic and normal gait. While the above-mentioned studies depend on using external sensors, Ortells et al. [40] explored automated video assessment to analyze Ataxic gait. However, the dataset used in the study was simulated by 10 healthy volunteers, so it did not contain real data from subjects with Ataxia.

2.4 Detecting Ataxia through Video Analysis

While automated video analysis is challenging, it has the potential to increase accessibility and reduce the subjectivity of care. It can offer preliminary screening or diagnosis for subjects that do not have access to clinical care. Automated video analysis can also help doctors by providing objective metrics, eventually reducing the subjectivity in assessments. Although there has been limited work on the video-based analysis of CA, the work of Jaroensri et al. [23] is notable. They introduced an automated method for quantifying motion impairment of Ataxia patients by analyzing video recordings of the Finger to Nose Test. They used a neural network for pose estimation and optical flow techniques to track participant hand motion. Their models performed comparably to

traditional clinical assessments of Ataxia severity. However, video-based analysis of CA for the gait task is not well-studied in the literature.

3   DATASET AND CLINICAL ANNOTATION

Table 1. Demographics information of our participants. PwA refers to People with Ataxia.

Status	#Participants (N=89)	Mean age (std dev)	#Female (percentage)	#Videos
PwA	65	47.07(9.08)	36 (55.38)	116
Control	24	41.73(8.72)	15 (62.5)	39
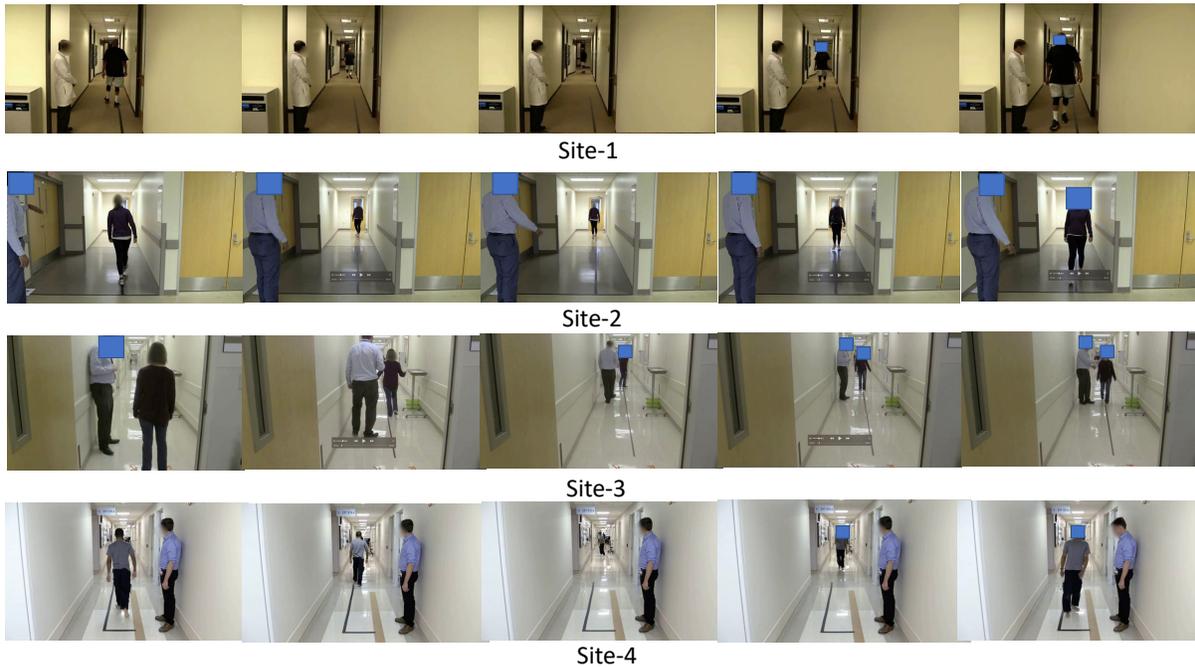
Fig. 2. Gait walk sample from four sites. We analyzed the first six seconds of data for each video. In all cases, the participants were still going away from the camera after six seconds (corresponding to the first two images for each site).

For this study, we collaborated with 11 different clinical sites across 8 different states in the United States and collected a dataset consisting of 183 recorded videos of participants completing the SARA gait task at their annual visits. During this task, a participant is supposed to walk through a corridor for approximately 10m, take a turn, and then walk back to the starting point. A neurologist observes the patient on the side and then assigns a score. This score works as the annotation for our supervised machine-learning approach. After the videos are captured, a separate clinician observes the video and marks the exact timestamp at the walking begins and ends.

We used the following exclusionary criteria to determine participant eligibility:

- Known genotype consistent with that of other inherited Ataxia (for example, Freidrich's Ataxia)
- Concomitant disorder(s) that affected the assessment of Ataxia severity during the study
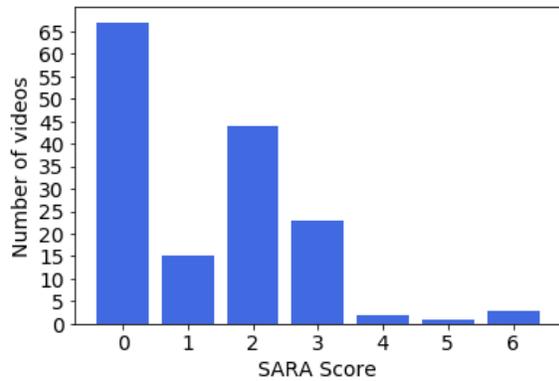
Fig. 3. Number of videos for each SARA score for gait
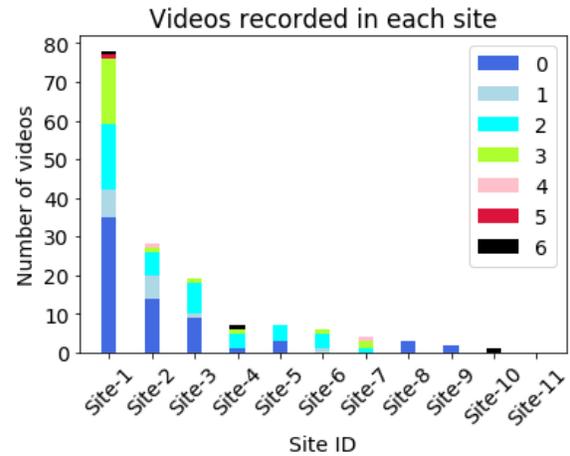
Fig. 4. Number of videos recorded at each site and gait-SARA score [0-6] mix in those videos
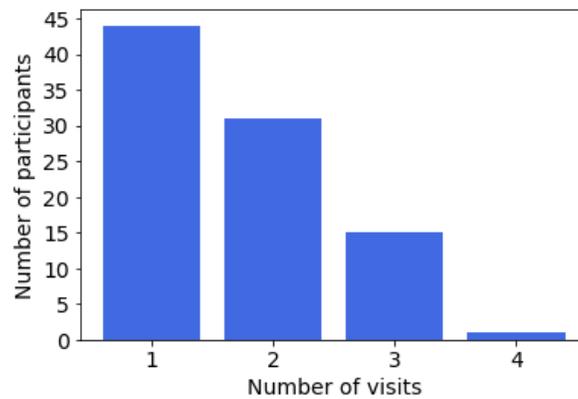
Fig. 5. Number of participants who completed 1, 2, 3, or 4 visits.

- Changes in physical and occupational Ataxia therapy in the two months prior to study enrollment

To reduce the possibility of identification, the participant's faces were blurred using Mediapipe Face Mesh[1] before transferring the data to our site. Upon removing noisy data (missing labels, interruption during walking, fully assisted walking, etc.), we were left with 155 videos from 89 participants. We collected 26, 66, 30, and 32 videos in 2018, 2019, 2020, and 2021 respectively. Among them, 24 were healthy controls and the remaining 65 were either already symptomatic with spinocerebellar Ataxia type 1 or type 3 (SCA1 and SCA3 respectively) or were otherwise pre-symptomatic. We will refer to these 65 participants as People-with-Ataxia (PwA). Across the 65 PwA and 24 controls, we obtained 116 and 39 unique videos respectively. The participants were diagnosed as either control or Ataxia patients at the beginning of the study and diagnosis has not changed for any participant.

---

[1]https://google.github.io/mediapipe/solutions/face_mesh.html

However, in some of the participants, the stage/degree of Ataxia may have changed which is reflected in the corresponding Gait scores for each video data. Table 1 provides demographic information for the participants.

According to the study protocol, both the control and PwA will provide data once every year for the duration of the study. Figure 5 shows that more than 50% of participants provided data more than once. However, since it is a multi-site longitudinal study, all sites did not start the study simultaneously. Besides, some participants dropped out of the study midway, and thus, we don't have the complete set of data for all participants.

Figure 4 provides an overview of the data count and the gait-specific SARA score distribution across sites. As can be seen, the first three sites provided most of the videos and have a good variety of videos with different SARA scores. Figure 2 provides snapshots of videos collected at the various locations. Figure 3 provides a breakdown of the distribution of gait-specific SARA scores across all 155 collected videos (PwA and control).

4 METHODS

Developing a supervised Machine Learning framework for predicting the Ataxia risk and severity from videos requires us to capture subtle differences in the participant's movement during the walk. Although the SARA evaluation metric is well established in the medical community, the task of predicting it from video data poses several challenges, for example,

(1) Each site has different lighting conditions, camera resolution, frame rate, and camera position, and uses different lengths of walking.
(2) The framework should be generalizable across different heights, ages, gender, weight, and other physical characteristics of the participants.
(3) Each video contains two or more people, i.e. the participant, one or more neurologists, and often time passersby. The framework has to separate participants from non-participants automatically.
(4) As the participants walk away from the camera, the camera perspective, and the participant's observed height change (Figure 2). The framework should be independent of these dynamic changes.

4.1 Dataset Preparation

We split the raw video with the start and end timestamp of the gait task provided by the clinicians. In order to eliminate the variability of different walking distances across sites, we only consider the first 6 seconds of the walking. During those 6 seconds, all participants were moving away from the camera and did not have time to turn around and return back towards the camera. We split these 6s videos into individual frames. With a frame rate of approximately 30 fps, every video is split into a sequence of around 180 images.

4.2 Person Detection and Tracking

The first step in analyzing the gait task from the video would be separating the participants from the background. In computer vision, this is a well-studied task known as Object Detection, and there are several open-source models for this task. We experimented with OpenCV Object detection [7], Yolo V4 [4], Yolo V5, and FasterRCNN [44] and found that FasterRCNN provided the most consistent and smooth boundaries for a person across frames. The object detection model takes an image frame as input and outputs a bounding box around each person in the frame.

However, the object detection model does not provide any information to track the person across frames. In order to do that, we applied the SORT algorithm [3] that takes the bounding boxes produced by FasterRCNN as input and assigned a unique ID to each person. Thus, we could generate a uniquely identifiable series of bounding boxes for all persons detected in the videos. This is presented in Figure 1, transition A to B.
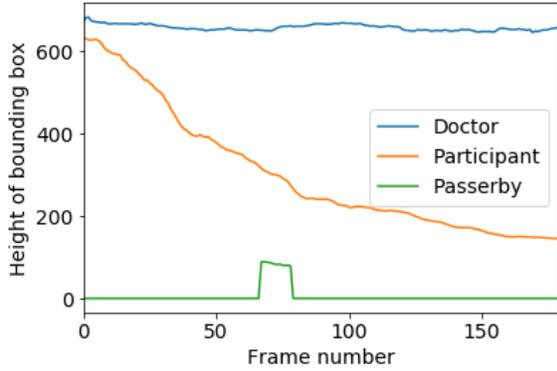
Fig. 6. Change of bounding box height for each identified person in a sample video (first 6s of the walk, 180 frames.)
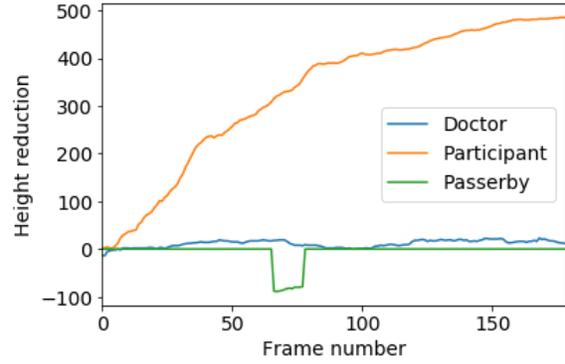
Fig. 7. *height_reduction* score (equation 1) for incremental frames for each identified person. Participant has the highest value–which helps separate the participant.

4.3 Gait Identification Algorithm

In the previous step, we assigned a unique ID for each person in a video. However, separating the participant from the non-participant is a non-trivial problem. There are different types of non-participants captured in the videos, for example, doctors who are evaluating the participant or passersby. The movement of non-participants is less predictable. To isolate the participant from non-participants, we develop a gait identification algorithm based on the nature of the gait task. A participant who is performing the gait task is expected to constantly move away from the camera and therefore consistently shrink in size (Figure 6). If we calculate the cumulative reduction of each person's height, the participant typically has the highest value (of this metric).

For each unique person identified by our object tracking algorithm, we calculate their cumulative height reduction for N frames with Equation 1. Here H_i represents the height of the bounding box for a person for ith frame.

$$height\_reduction = \sum_{i=0}^{N-1} (H_i - H_{i+1}) \tag{1}$$

We calculated a *height_reduction* for each of the persons tracked; the individual with the highest score has walked the farthest distance consistently, hence, they are the participant (Figure 6). As doctors typically show relatively less forward movement than the participant, and passersby usually appear for fewer number of frames, they automatically get a lower *height_reduction* score. If the top two *height_reduction* scores in a video are within 90% of each other, it indicates an assisted walk, where the patient's condition is so severe that they need assistance from the doctor to complete the gait task. We discard these videos from our dataset, as our goal is to identify non-trivial cases of SCA. Finally, we manually observe all outputs and make sure that our method can successfully detect the participant in all videos

4.4 Gait Pose Estimation

In computer vision, pose estimation is the task of using a Machine Learning model to estimate the pose of a person from an image by calculating the spatial locations of key body joints (a.k.a key points). Using pose estimation models we can precisely track different key points of a participant's body, and use this for classifying

Table 2. Description of features. Since we initially extract a sequence of features for each Feature group entry (except `height_reduction`), we calculate six sub-features to represent that sequence: mean, std-deviation, max, min, range (max-min), and entropy. SL and SW represent step length and step width respectively. The values 1 and 2 in *Source* column represent [8] and [20] respectively.

Feature group	Measurement	Estimation for	Source
height_reduction	Total height reduction for the 6s walk (equation 1)	Speed	1
feet_dist	Distance between feet	SL, SW, swing	1, 2
feet_angle	Angle between the x-axis and line joining feet	Swing, SL	1
right_x	X-Coordinate of right feet	SL, SW, swing	1, 2
right_y	Y-Coordinate of right feet	SL, SW, swing	1, 2
left_x	X-Coordinate of left feet	SL, SW, swing	1, 2
left_y	Y-Coordinate of left feet	SL, SW, swing	1, 2
left_knee_bent	Angle between the left thigh and knee	Swing	1
right_knee_bent	Angle between the right thigh and knee	Swing	1
imbalance	Horizontal deviation of the mid-shoulders and hip	Stability	2
tilt	Angle of shoulder line with horizontal axis	Stability	2
nose_x	X-coordinate of the nose	Stability	2
nose_y	Y-coordinate of the nose	Stability	2
stance	% of frames where the participant was stationary	Stance phase	1

the Ataxia risk. We experimented with several open-source pose estimation models, and we found MoveNet[2] in the Tensorflow library to provide consistent pose estimation.

Once the participant is identified in a video in the previous section, we crop the participant's bounding box from each frame and upscale or downscale them to uniform 800×400 pixel images. This scaling eliminates the variation in the dataset regarding participants' height, distance from the camera, camera resolution, and perspective. From each one of these images, we use MoveNet to detect 17 key points of the participant's body–which is later used to calculate meaningful features for our prediction task.

4.5 Feature Extraction from Pose

After separating out the frames containing the participant of interest, we extracted several features based on previously validated gait characteristics of SCA patients [8]. Notably, these features reflect common Ataxia gait features including step width, step length, swing, stability, stance phase, swing phase, and speed.

Table 2 provides a short description of our features. Studies found reduced walking speed as an important characteristic of SCA patients [8]. For our case in a fixed time window (6 seconds), walking speed will translate to the furthest distance covered–which is indirectly captured by our `height_reduction` score calculated in Section 4.3 (person who travels further, would get smaller quickly). We use `height_reduction` score calculated for each patient as a feature for the classifier. The next 13 entries in Table 2 represent discrete signals, each consisting of nearly 180 timestamps. In order to capture learnable parameters from these signals, we identified six aggregate quantities: mean, std-deviation, max, min, range ($max - min$), and entropy. The entropy of a signal measures the irregularity in the data. Mathematically, a periodic signal will have zero entropy, and a random signal will have high entropy. As SCA patients have difficulties maintaining a consistent rhythm in their walking, entropy-related features should effectively capture these irregularities in a walking pattern. We used an open-source Python

---

[2]https://www.tensorflow.org/hub/tutorials/movenet

implementation [56] of the spectral entropy method originally designed to capture irregularities in EEG brain signals [21].

Thus, we finally have $79(1 + 13 * 6 = 79)$ features to build our models from which we select the final features using Recursive Feature Elimination with Cross-Validation (RFECV) algorithm.

4.6 Ataxia Risk-Prediction

Our first goal in this study (RQ1) is to automatically recognize an individual's risk of having SCA. In order to do that with the help of our annotated dataset, we label the participants with a SARA gait score of 0 as 'risk-free' (Class 0) and participants with a score > 0 score as 'at risk' (Class 1) and train a Machine Learning classifier to classify between these two. We employed the recursive feature elimination with cross-validation to select the optimal set of features and train a Random Forest model as the classifier. The complete list of features used for training this model is available in A.1. We evaluated the model using 10-fold cross-validation and leave-one-site-out methodology. In the second case, we trained the model on all sites but one and tested the model on the held-out site. Good performance on this task provides a useful metric for evaluating our model's generalizability in completely unseen situations.

There are 88 videos in Class 1 and 67 videos in Class 0. Therefore, the accuracy of choosing the majority class is 0.568. Hence, an accuracy of more than 0.568 indicates the model is performing better than random chance. We perform all of the experiments 20 times and average the results to minimize randomness.

To evaluate how our model performs on PwA with subtle symptoms, we ran the same experiment by removing all videos with SARA Gait score > 3. Therefore, gait score 0 is in Class 0, and gait score 1, 2, 3 is in Class 1. From Table 4, we can see that except for a 7% decrease in accuracy for the "Site 2" sub-study, the performance remains on par with those reported previously in Table 3. However, since we are working with two different datasets (due to removing videos with gait scores > 3), we should be cautious about making a direct comparison between these two experiments.

4.7 Ataxia Severity-assessment

Our second goal (RQ2) is to predict the SARA score directly from the gait data, which indicates the severity of the disease. We design this task as a regression problem and use a random forest regression model to predict a continuous SARA score value from the video-extracted data. As our primary motivation is to identify early-stage CA, we focus on lower scores of SARA and categorize all SARA scores of 3 or above to simply 3. The model takes the numeric features from participants' walking as input and predicts a continuous score between 0 and 3. We use the feature importance score of the random forest algorithm to identify the best features for training this model (a complete list is available in A.2. We evaluate the regression model by calculating the mean absolute error (MAE) and Pearson's correlation coefficient (PCC).

To predict severity for PwA with mild symptoms, we repeated the above experiment by removing videos with gait SARA score > 3 and reported the performance in Table 6. Although the models' performances degrade slightly for '10-fold CV', 'Site 5', and 'Site 6', it improves or remains on par with those reported in Table 6.

5 RESULTS AND INTERPRETATION

5.1 Ataxia Risk-prediction Results

The mean and standard deviation of the 20 runs of the Ataxia Risk-prediction are reported in Table 3. We can see from Table 3 that the model achieves a mean F1 score of 80.23% and an accuracy of 83.06% in 10-fold cross-validation. Clinically, the model is able to distinguish a CA patient from a non-CA subject with above 80% correctness. The low standard deviation between results indicates that the results are fairly consistent

Table 3. Ataxia risk prediction (binary classification) model performance. The model was evaluated in 20 iterations of 10-fold cross-validation and leave-one-site-out manner (Section 4.6).

Test	F1 score (%)		Accuracy (%)	
	Mean	STD	Mean	STD
10-fold CV	80.23	9.19	83.06	6.79
Site 2	80.5	2.65	80.54	2.64
Site 3	65.34	5.04	65.79	4.85
Site 4	74.65	6.32	81.43	6.55
Site 5	66.94	5.83	72.86	4.29
Site 6	94.55	16.36	98.33	5.0

Table 4. Ataxia risk prediction (binary classification, excluding all data with label 4,5,6) model performance. The model was evaluated in 20 iterations of 10-fold cross-validation and leave-one-site-out manner (Section 4.6).

Test	F1 score (%)		Accuracy (%)	
	Mean	STD	Mean	STD
10-fold CV	80.21	9.89	81.78	9.32
Site 2	73.07	3.31	73.15	3.29
Site 3	70.8	8.69	71.05	8.57
Site 4	77.01	3.33	82.5	3.63
Site 5	69.86	8.42	75.0	6.19
Site 6	100.0	0.0	100.0	0.0

Table 5. Ataxia severity prediction (regression between 0~3) model performance. The model is evaluated with 20 iterations of 10-fold cross-validation and a leave-one-site-out manner. In the second case, a model is tested on a site that was excluded during training (Section 4.7). The mean and standard deviation of Mean Absolute Error (MAE) (lower is better) and Pearson's Correlation Coefficient (higher is better) is reported.

Test	MAE		Corr Coeff.	
	Mean	STD	Mean	STD
10-fold CV	0.6225	0.0132	0.7268	0.0144
Site 2	0.5673	0.0	0.7081	0.0
Site 3	0.7723	0.0	0.4569	0.0
Site 4	1.0886	0.0	0.2146	0.0
Site 5	0.6334	0.0	0.6244	0.0
Site 6	0.4402	0.0	0.4297	0.0

between different runs. In leave-one-site-out tests, the model performs fairly well, despite never seeing data from a particular site.

5.2 Ataxia Severity-assessment Results

On 10-fold-cross-validation, our model achieves MAE and PCC scores of 0.6225 and 0.7268 respectively. A lower MAE is better, as it indicates the mean difference between the model's prediction and the actual SARA score. PCCs range from -1 to 1, where higher values are better, implying a correlation between the prediction and

Table 6. Ataxia severity prediction (regression between 0~3 excluding all data with score 4 and above) model performance. The model is evaluated with 20 iterations of 10-fold cross-validation and a leave-one-site-out manner. In the second case, a model is tested on a site that was excluded during training (Section 4.7). The mean and standard deviation of Mean Absolute Error (MAE) (lower is better) and Pearson's Correlation Coefficient (higher is better) is reported.

Test	MAE		Corr Coeff.	
	Mean	STD	Mean	STD
10-fold CV	0.7035	0.0135	0.6885	0.0126
Site 2	0.5644	0.0	0.5738	0.0
Site 3	0.72	0.0	0.6192	0.0
Site 4	0.7187	0.0	0.5796	0.0
Site 5	0.7565	0.0	0.3549	0.0
Site 6	0.5197	0.0	0.0608	0.0

ground truth. We can see from Table 5 that in most cases, the model achieves less than 1.0 MAE through the leave-one-site analysis and its predictions are positively correlated with ground truth SARA scores. The low standard deviation indicates that the predictions are consistent across runs. By manually inspecting 6 data points (Figure 2) in Site 6, we find that although the model has high accuracy in Table 3 and the lowest of all MAE in Table 5, a single misclassification results in a relatively lower but still acceptable correlation score. Figures 9 (a) and 9 (b) visualize the positive correlation through a scatter plot and Gaussian kernel density estimation.

5.3 Model Interpretation

Figure 8 (a) and 8 (b) give us insights into which features have a greater influence on our models' predictions. We can see that the coordinates of the left and right feet, as well as the distance of steps, play important roles in model prediction. Based on these summaries, we see that wider steps (min_left_x, max_feet_dist, max_right_x, etc.) and variation in steps (std_left_x, range_feet_dist, etc.) both affect Ataxia scores; this is consistent with [8, 45, 57], who also found that step length, step width, step distance, and differences in step length between the left and right feet are indicative of CA. [8, 45] also reported that individuals with CA have lower walking speeds than healthy controls, which is captured by the height_reduction score. We can see that in both graphs, height_reduction negatively affects the CA score. In addition, [45] found that Ataxia participants' have oscillatory walking patterns, which are effectively captured through the imbalance features in our model; range_imbalance, std_imbalance, min_imbalance, etc.) contribute to higher SARA scores, thus denoting more advanced Ataxia. Position of the nose (Figure 8 (b)) can also be indicative of oscillation in upper body movement. Overall, Figure 8 (a) and 8 (b) demonstrate that the most significant features in our model are consistent with those identified in prior literature.

6 LIMITATIONS

6.1 Study Design and Feature-set

In this study, the classification features are computed with clever engineering grounded on the existing literature. However, a more exhaustive study of Ataxia literature and its effects could reveal more discriminative features that can result in better performance from our dataset. In order to normalize different walking distances between sites, we only considered the first 6 seconds of walking forward part of the video-recorded gait. However, in a standard setting, a patient walks forward, turns, and then walks back to the initial position. Taking the full walk into consideration could potentially provide more information.

(a) (a) Risk prediction model (binary)   (b) (b) Severity assessment model (regression)
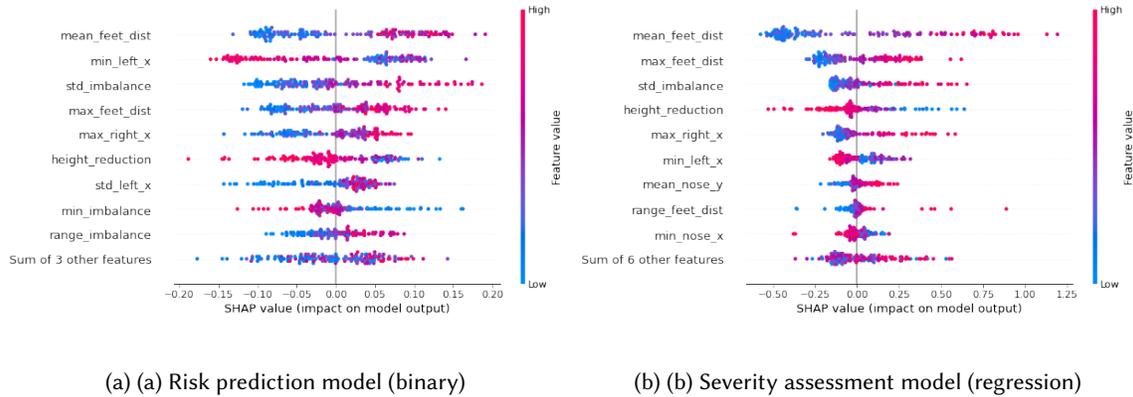
Fig. 8. Features' impact on models prediction (computed by SHAP([33])). Each dot in the figure represents the impact of that feature in one data point. The dots' colors are arranged in a blue-to-red spectrum – blue indicates the lower value of that feature and red indicates higher. Similarly, where that dot is placed on the x-axis indicates how it's impacting the model's output. If the feature is pushing the output towards higher values, it is placed right to the vertical line at zero. If it is pushing towards lower output values, it is placed left that line. For example, in both sub-figures, higher values of mean_feet_dist are increasing the risk probability and severity value respectively; this is intuitive since Ataxia patients tend to have a wider base (distance between feet) while walking.

(a) (a) Scatter plot   (b) (b) Gaussian kernel density estimation
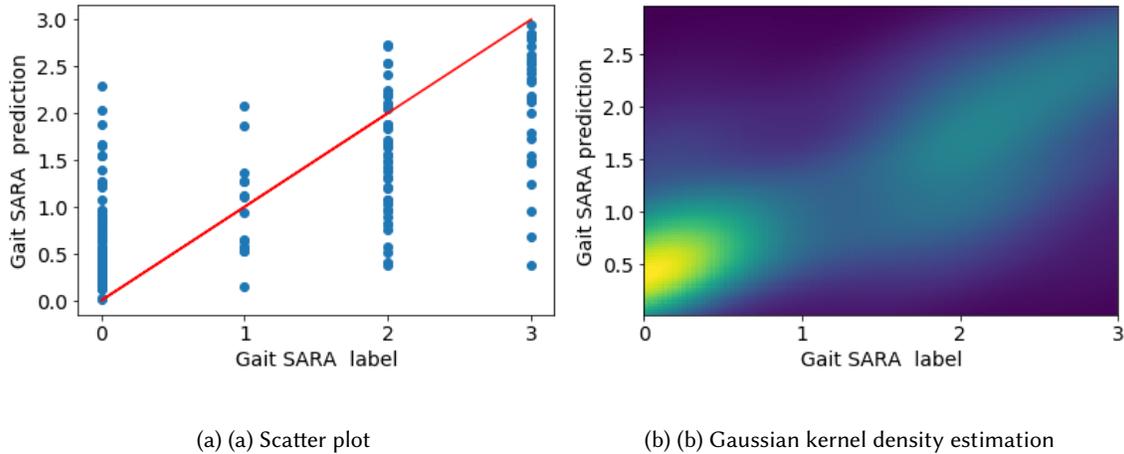
Fig. 9. Visualizing the predictions for Severity-assessment task. The red line in Fig. *(a)* denotes the ideal scatter plot. The light-green color in Fig. *(b)* shows an estimation of the prediction density. The lighter the color, the higher the number of predictions in that region.

With our feature analysis techniques, we were able to computationally explain the different features used for classification with findings from the clinical literature. Our dataset can potentially reveal more such findings about the characteristic of Cerebellar Ataxia (SCA) than what we have been able to demonstrate in the paper. By opening the dataset to the research community, we hope that it will develop novel computational models and interpretable results for detecting Cerebellar Ataxia.

Due to dataset size and the desire to have transparent and interpretable metrics, our effort to achieve high accuracy using the more advanced deep learning techniques did not pan out. Incorporating the latest techniques into our algorithmic workflow remains part of our future work.

6.2 Generalizability

Our dataset was collected on people either diagnosed or at risk of Cerebellar Ataxia (SCA). While SARA can be also used to assess other types of Ataxia aside from SCA – such as Ataxia due to brain tumors and Ataxia telangiectasia [41] – the target populations nonetheless differ. Therefore, our results may not translate to other types of Ataxia. In addition, the videos were collected in clinical settings, and all the participants were assisted by a doctor. It would be interesting to explore whether our findings hold for home-recorded videos, potentially without any supervision of clinical personnel.

7 FUTURE WORK

7.1 Home Applications

Our broader methodology of detecting, separating, and tracking a clinical participant and predicting their disease diagnosis by tracking body poses has several clinical applications. With minor modification, our methodology is applicable to Tandem Gait – another task in SARA evaluation. Nonetheless, this work opens up possibilities for the development of computer vision frameworks for more tasks on SARA evaluation. Our video-based analysis method is robust to different environments, backgrounds, and the number of people present in a video frame. Our feature set is also based on movement and independent of the participant's age, gender, height, weight, and other physical characteristics. This makes our model suitable to be used outside controlled clinical settings and possibly at home setup [17]. Deployment of such a diagnosis tool at home setup will make neurological care accessible for a large population and provide earlier interventions to vulnerable people.

7.2 Differentiating between Different Movement Disorders

Gait is a popular diagnosis task for a different types of movement disorders. For example, both SARA for measuring ataxia severity and the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) for measuring Parkinson's include a gait task [16, 51]. Different movement disorders often present similar types of symptoms which results in clinical misdiagnosis. In the words of an Ataxia patient, *"I know there are people walking around with exactly what I have and they don't know it.... It is often misdiagnosed because it resembles other things like multiple sclerosis, stroke, and Parkinson's disease."*[5]. Our pipeline and dataset can further be used to distinguish between different movement disorders and help reduce misdiagnosis.

7.3 Human-centric Applications

In the future, we plan to expand our work in several directions. For example, we can design systems to help doctors annotate relevant parts in the video that contain important information for detecting Ataxia [10]. Similarly, we can integrate multimodal data – gait task, demographic information, stance task, etc – to better help the neurologists [11]. Moreover, if we want to integrate our tool in the Ataxia diagnosis pipeline, then we will need to build a proper human-centric AI assistant [12] and quantify how it helps neurologists better diagnose Ataxia [13].

8 CONCLUSION

Ataxia is a rare disorder and its treatment is only available in urban hospitals for which an in-person visit is necessary for diagnosis and continuous treatment of the condition. Although sensor-based systems have shown

promising results in diagnosing Ataxia in lab settings, equipping people with sensors or installing high-end 3D cameras is not scalable, which is why algorithms that work on the existing and diverse infrastructure are needed. In this paper, we present an end-to-end computer vision pipeline that can detect participants displaying Ataxia-impacted gait characteristics and can measure the severity of Ataxia from the SARA gait task. Our models open up new possibilities for the future development of a user-friendly, automated platform that can remotely assess Ataxia in non-clinical settings. This would have various applications, including independent use by patients and working as an objective clinical outcome measure for pharmaceutical companies, academic institutions, and other key stakeholders. Our technology could aid clinicians in having a second opinion in their existing diagnosis pipeline. Besides our general methodology of person detection, tracking, and separation, it can be further utilized to identify gait anomalies in elderly individuals who are vulnerable to falls [29], which is one of the main reasons for fractures, physical incapacity, and even death [54] in the elderly population. Our dataset will help the computer science community to develop better algorithms to quantify and potentially diagnose movement disorders. Our method is scalable and interpretable, which is helpful for future deployment and possible adoption in healthcare settings.

9 INSTITUTIONAL REVIEW BOARD (IRB)

We conducted the experiments with IRB approval of the relevant institution.

10 DATASET AND CODE AVAILABILITY

To avoid potential malicious use of our dataset (e.g. surveillance/stalking of people with Ataxia), we will provide the anonymized dataset consisting of extracted pose features and SARA score on a case-by-case request basis upon request. Each request will be evaluated by IRB for approval. This anonymized dataset will contain 17 key data points per image from each video (180 images per video). Furthermore, Willing collaborators can be added to the protocol with full data access based on approval from the IRB. The code used in this project and further instructions to receive the data are available at https://github.com/Masum06/Automated-Ataxia-Gait.

ACKNOWLEDGMENTS

REFERENCES

[1] Kamiar Aminian, Farzin Dadashi, Benoit Mariani, Constanze Lenoble-Hoskovec, Brigitte Santos-Eggimann, and Christophe J. Büla. 2014. Gait Analysis Using Shoe-Worn Inertial Sensors: How is Foot Clearance Related to Walking Speed?. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14).* Association for Computing Machinery, New York, NY, USA, 481–485. https://doi.org/10.1145/2632048.2632071

[2] Boyd Anderson, Mingqian Shi, Vincent Y. F. Tan, and Ye Wang. 2019. Mobile Gait Analysis Using Foot-Mounted UWB Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 73 (sep 2019), 22 pages. https://doi.org/10.1145/3351231

[3] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. 2016. Simple online and realtime tracking. In *2016 IEEE international conference on image processing (ICIP).* IEEE, 3464–3468.

[4] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934* (2020).

[5] Body and Mind staff. 2021. Ataxia sufferer faces a neurological disorder that's often misdiagnosed. https://www.pennlive.com/bodyandmind/2011/11/positive_attitude_is_best_medi.html.

[6] Brian M Bot, Christine Suver, Elias Chaibub Neto, Michael Kellen, Arno Klein, Christopher Bare, Megan Doerr, Abhishek Pratap, John Wilbanks, E Dorsey, et al. 2016. The mPower study, Parkinson disease mobile data collected using ResearchKit. *Scientific data* 3, 1 (2016), 1–9.

[7] Gary Bradski and Adrian Kaehler. 2000. OpenCV. *Dr. Dobb's journal of software tools* 3 (2000), 2.

[8] Ellen Buckley, Claudia Mazzà, and Alisdair McNeill. 2018. A systematic review of the gait characteristics associated with Cerebellar Ataxia. *Gait & posture* 60 (2018), 154–163.

[9] Katrin Bürk, Ulrike Mälzig, Stefanie Wolf, Suzette Heck, Konstantinos Dimitriadis, Tanja Schmitz-Hübsch, Sascha Hering, Tobias M Lindig, Verena Haug, Dagmar Timmann, et al. 2009. Comparison of three clinical rating scales in Friedreich ataxia (FRDA). *Movement disorders* 24, 12 (2009), 1779–1784.

[10] Francisco M Calisto, Alfredo Ferreira, Jacinto C Nascimento, and Daniel Gonçalves. 2017. Towards touch-based medical image diagnosis annotation. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. 390–395.

[11] Francisco Maria Calisto, Nuno Nunes, and Jacinto C Nascimento. 2020. BreastScreening: on the use of multi-modality in medical imaging diagnosis. In *Proceedings of the international conference on advanced visual interfaces*. 1–5.

[12] Francisco Maria Calisto, Carlos Santiago, Nuno Nunes, and Jacinto C Nascimento. 2021. Introduction of human-centric AI assistant to aid radiologists for multimodal breast image classification. *International Journal of Human-Computer Studies* 150 (2021), 102607.

[13] Francisco Maria Calisto, Carlos Santiago, Nuno Nunes, and Jacinto C Nascimento. 2022. BreastScreening-AI: Evaluating medical intelligent agents for human-AI interactions. *Artificial Intelligence in Medicine* 127 (2022), 102285.

[14] E Ray Dorsey, Bastiaan R Bloem, and Michael S Okun. 2020. A new day: the role of telemedicine in reshaping care for persons with movement disorders. *Movement Disorders* 35, 11 (2020), 1897–1902.

[15] Ondřej Dostál, Aleš Procházka, Oldřich Vyšata, Ondřej Ťupa, Pavel Cejnar, and Martin Vališ. 2021. Recognition of motion patterns using accelerometers for ataxic gait assessment. *Neural Computing and Applications* 33, 7 (2021), 2207–2215.

[16] Christopher G Goetz, Barbara C Tilley, Stephanie R Shaftman, Glenn T Stebbins, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Matthew B Stern, Richard Dodel, et al. 2008. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Movement disorders: official journal of the Movement Disorder Society* 23, 15 (2008), 2129–2170.

[17] Marcus Grobe-Einsler, Arian Taheri Amin, Jennifer Faber, Tamara Schaprian, Heike Jacobi, Tanja Schmitz-Hübsch, Alhassane Diallo, Sophie Tezenas du Montcel, and Thomas Klockgether. 2021. Development of SARAhome, a New Video-Based Tool for the Assessment of Ataxia at Home. *Movement Disorders* 36, 5 (2021), 1242–1246.

[18] Michael Hardegger, Gerhard Tröster, and Daniel Roggen. 2013. Improved ActionSLAM for Long-Term Indoor Tracking with Wearable Motion Sensors. In *Proceedings of the 2013 International Symposium on Wearable Computers* (Zurich, Switzerland) *(ISWC '13)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/2493988.2494328

[19] H Hartley, B Pizer, S Lane, C Sneade, R Pratt, A Bishop, and R Kumar. 2015. Inter-rater reliability and validity of two ataxia rating scales in children with brain tumours. *Child's Nervous System* 31, 5 (2015), 693–697.

[20] Takeru Honda, Hiroshi Mitoma, Hirotaka Yoshida, Kyota Bando, Hiroo Terashi, Takeshi Taguchi, Yohane Miyata, Satoko Kumada, Takashi Hanakawa, Hitoshi Aizawa, et al. 2020. Assessment and rating of motor cerebellar ataxias with the Kinect v2 depth sensor: extending our appraisal. *Frontiers in neurology* 11 (2020), 179.

[21] Tsuyoshi Inouye, Kazuhiro Shinosaki, H Sakamoto, Seigo Toi, Satoshi Ukai, Akinori Iyama, Y Katsuda, and M Hirano. 1991. Quantification of EEG irregularity by use of the entropy of the power spectrum. *Electroencephalography and clinical neurophysiology* 79, 3 (1991), 204–210.

[22] Heike Jacobi, Till-Karsten Hauser, Paola Giunti, Christoph Globas, Peter Bauer, Tanja Schmitz-Hübsch, László Baliko, Alessandro Filla, Caterina Mariotti, Maria Rakowicz, et al. 2012. Spinocerebellar ataxia types 1, 2, 3 and 6: the clinical spectrum of ataxia and morphometric brainstem and cerebellar findings. *The Cerebellum* 11, 1 (2012), 155–166.

[23] Ronnachai Jaroensri, Amy Zhao, Guha Balakrishnan, Derek Lo, Jeremy D Schmahmann, Frédo Durand, and John Guttag. 2017. A video-based method for automatically rating ataxia. In *Machine Learning for Healthcare Conference*. PMLR, 204–216.

[24] Jeyeon Jo and Huiju Park. 2021. RFInsole: Batteryless Gait-Monitoring Smart Insole Based on Passive RFID Tags. In *2021 International Symposium on Wearable Computers* (Virtual, USA) *(ISWC '21)*. Association for Computing Machinery, New York, NY, USA, 141–143. https://doi.org/10.1145/3460421.3478810

[25] Hsin-Liu (Cindy) Kao, Bo-Jhang Ho, Allan C. Lin, and Hao-Hua Chu. 2012. Phone-Based Gait Analysis to Detect Alcohol Usage. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. Association for Computing Machinery, New York, NY, USA, 661–662. https://doi.org/10.1145/2370216.2370354

[26] Mustafa Kaya, Serkan Karakuş, and Seda Arslan Tuncer. 2022. Detection of ataxia with hybrid convolutional neural network using static plantar pressure distribution model in patients with multiple sclerosis. *Computer Methods and Programs in Biomedicine* 214 (2022), 106525.

[27] Navleen Kour, Sunanda, and Sakshi Arora. 2019. Computer-Vision Based Diagnosis of Parkinson's Disease via Gait: A Survey. *IEEE Access* 7 (2019), 156620–156645. https://doi.org/10.1109/ACCESS.2019.2949744

[28] Ragil Krishna, Pubudu N Pathirana, Malcolm Horne, Laura Power, and David J Szmulewicz. 2019. Quantitative assessment of cerebellar ataxia, through automated limb functional tests. *Journal of neuroengineering and rehabilitation* 16, 1 (2019), 1–15.

[29] Ingebjørg Lavrantsdatter Kyrdalen, Pernille Thingstad, Leiv Sandvik, and Heidi Ormstad. 2019. Associations between gait speed and well-known fall risk factors among community-dwelling older adults. *Physiotherapy research international* 24, 1 (2019), e1743.

[30] Raina Langevin, Mohammad Rafayet Ali, Taylan Sen, Christopher Snyder, Taylor Myers, E Ray Dorsey, and Mohammed Ehsan Hoque. 2019. The PARK Framework for Automated Analysis of Parkinson's Disease Characteristics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 1–22.

[31] Robert LeMoyne, Frederic Heerinckx, Tanya Aranca, Robert De Jager, Theresa Zesiewicz, and Harry J Saal. 2016. Wearable body and wireless inertial sensors for machine learning classification of gait for people with Friedreich's ataxia. In *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 147–151.

[32] Jiarong Li, Zihan Wang, Zihao Zhao, Yuchao Jin, Jihong Yin, Shao-Lun Huang, and Jiyu Wang. 2021. TriboGait: A Deep Learning Enabled Triboelectric Gait Sensor System for Human Activity Recognition and Individual Identification. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers* (Virtual, USA) *(UbiComp '21)*. Association for Computing Machinery, New York, NY, USA, 643–648. https://doi.org/10.1145/3460418.3480410

[33] Scott M Lundberg, Bala Nair, Monica S Vavilala, Mayumi Horibe, Michael J Eisses, Trevor Adams, David E Liston, Daniel King-Wai Low, Shu-Fang Newman, Jerry Kim, et al. 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering* 2, 10 (2018), 749.

[34] Mario Manto, Nicolas Dupre, Marios Hadjivassiliou, Elan D Louis, Hiroshi Mitoma, Marco Molinari, Aasef G Shaikh, Bing-Wen Soong, Michael Strupp, Frank Van Overwalle, et al. 2020. Medical and paramedical care of patients with cerebellar Ataxia during the COVID-19 outbreak: seven practical recommendations of the COVID 19 cerebellum task force. *Frontiers in neurology* 11 (2020), 516.

[35] Mayo-Clinic-Stuff. 2021. Ataxia. https://www.mayoclinic.org/diseases-conditions/ataxia/symptoms-causes/syc-20355652.

[36] Mostafa Mirshekari, Jonathon Fagert, Amelie Bonde, Pei Zhang, and Hae Young Noh. 2018. Human Gait Monitoring Using Footstep-Induced Floor Vibrations Across Different Structures. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (Singapore, Singapore) *(UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 1382–1391. https://doi.org/10.1145/3267305.3274187

[37] Khoa D Nguyen, Pubudu N Pathirana, Malcolm Horne, Laura Power, and David J Szmulewicz. 2020. Entropy-based analysis of rhythmic tapping for the quantitative assessment of cerebellar ataxia. *Biomedical Signal Processing and Control* 59 (2020), 101916.

[38] Arinobu Niijima, Kazuhiro Yoshida, Osamu Mizuno, Yumiko Tanmatsu, Naoki Asanoma, Tomoki Watanabe, Tsubasa Nakayama, and Makoto Oyama. 2016. Estimation of Beautiful Gait Using an Accelerometer. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 169–172. https://doi.org/10.1145/2968219.2971408

[39] Javier Ortells, María Trinidad Herrero-Ezquerro, and Ramón A Mollineda. 2018. Vision-based gait impairment analysis for aided diagnosis. *Med Biol Eng Comput.* (feb 2018). https://doi.org/10.1007/s11517-018-1795-2

[40] Javier Ortells, María Trinidad Herrero-Ezquerro, and Ramón A Mollineda. 2018. Vision-based gait impairment analysis for aided diagnosis. *Medical & biological engineering & computing* 56, 9 (2018), 1553–1564.

[41] Santiago Perez-Lloret, Bart Van de Warrenburg, Malco Rossi, Carmen Rodríguez-Blázquez, Theresa Zesiewicz, Jonas AM Saute, Alexandra Durr, Masatoyo Nishizawa, Pablo Martinez-Martin, Glenn T Stebbins, et al. 2021. Assessment of ataxia rating scales and cerebellar functional tests: critique and recommendations. *Movement Disorders* 36, 2 (2021), 283–297.

[42] Dung Phan, Nhan Nguyen, Pubudu N Pathirana, Malcolm Horne, Laura Power, and David Szmulewicz. 2019. A random forest approach for quantifying gait ataxia with truncal and peripheral measurements using multiple wearable sensors. *IEEE Sensors Journal* 20, 2 (2019), 723–734.

[43] Aleš Procházka, Ondřej Dostál, Pavel Cejnar, Hagar Ibrahim Mohamed, Zbyšek Pavelek, Martin Vališ, and Oldřich Vyšata. 2021. Deep Learning for Accelerometric Data Assessment and Ataxic Gait Monitoring. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29 (2021), 360–367.

[44] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* 28 (2015), 91–99.

[45] T1 Schmitz-Hübsch, S Tezenas Du Montcel, L Baliko, J Berciano, S Boesch, Chantal Depondt, P Giunti, C Globas, J Infante, J-S Kang, et al. 2006. Scale for the assessment and rating of ataxia: development of a new clinical scale. *Neurology* 66, 11 (2006), 1717–1720.

[46] KJ Schouwstra, SS Polet, S Hbrahimgel, AS Tadema, JGM Burgerhof, R Brandsma, and DA Sival. 2022. Application of the Scale for Assessment and Rating of Ataxia in toddlers. *European Journal of Paediatric Neurology* 40 (2022), 28–33.

[47] Jill Seladi-Schulman. 2020. Ataxia: Definition, types, causes, diagnosis, treatment. https://www.healthline.com/health/ataxia#types-of-ataxia

[48] Mariano Serrao and Carmela Conte. 2018. Detecting and measuring ataxia in gait. In *Handbook of Human Motion*. Springer International Publishing, 937–954.

[49] Cosmin Stamate, George D Magoulas, Stefan Küppers, Effrosyni Nomikou, Ioannis Daskalopoulos, Ashwani Jha, JS Pons, J Rothwell, Marco U Luchini, Theano Moussouri, et al. 2018. The cloudUPDRS app: A medical device for the clinical assessment of Parkinson's Disease. *Pervasive and mobile computing* 43 (2018), 146–166.

[50] Cosmin Stamate, George D Magoulas, Stefan Küppers, Effrosyni Nomikou, Ioannis Daskalopoulos, Marco U Luchini, Theano Moussouri, and George Roussos. 2017. Deep learning Parkinson's from smartphone data. In *2017 IEEE international conference on pervasive computing and communications (PerCom)*. IEEE, 31–40.

[51] Sub H Subramony. 2007. SARA—a new clinical scale for the assessment and rating of ataxia. *Nature clinical practice Neurology* 3, 3 (2007), 136–137.

[52] Susanna Summa, Tommaso Schirinzi, Giuseppe Massimo Bernava, Alberto Romano, Martina Favetta, Enza Maria Valente, Enrico Bertini, Enrico Castelli, Maurizio Petrarca, Giovanni Pioggia, et al. 2020. Development of SaraHome: a novel, well-accepted, technology-based assessment tool for patients with ataxia. *Computer methods and programs in biomedicine* 188 (2020), 105257.

[53] S Summa, G Tartarisco, M Favetta, A Buzachis, A Romano, GM Bernava, G Vasco, G Pioggia, M Petrarca, E Castelli, et al. 2020. Spatio-temporal parameters of ataxia gait dataset obtained with the Kinect. *Data in Brief* 32 (2020), 106307.

[54] Miguel Terroso, Natacha Rosa, Antonio Torres Marques, and Ricardo Simoes. 2014. Physical consequences of falls in the elderly: a literature review from 1995 to 2010. *European Review of Aging and Physical Activity* 11, 1 (2014), 51–59.

[55] Ha Tran, Pubudu N Pathirana, Malcolm Horne, Laura Power, and David J Szmulewicz. 2019. Automated Evaluation of Upper Limb Motor Impairment of Patient with Cerebellar Ataxia. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 6846–6849.

[56] Rafayel Vallat. 2022. entropy.spectral_entropy. https://raphaelvallat.com/entropy/build/html/generated/entropy.spectral_entropy.html.

[57] Oldřich Vyšata, Ondřej Ťupa, Aleš Procházka, Rafael Doležal, Pavel Cejnar, Aprajita Milind Bhorkar, Ondřej Dostál, and Martin Vališ. 2021. Classification of Ataxic Gait. *Sensors* 21, 16 (2021). https://doi.org/10.3390/s21165576

[58] Wei Wang, Alex X. Liu, and Muhammad Shahzad. 2016. Gait Recognition Using Wifi Signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 363–373. https://doi.org/10.1145/2971648.2971670

[59] Wei Xu, ZhiWen Yu, Zhu Wang, Bin Guo, and Qi Han. 2019. AcousticID: Gait-Based Human Identification Using Acoustic Signal. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 115 (sep 2019), 25 pages. https://doi.org/10.1145/3351273

[60] Yang Xu, Wei Yang, Jianxin Wang, Xing Zhou, Hong Li, and Liusheng Huang. 2018. WiStep: Device-Free Step Counting with WiFi Signals. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 172 (jan 2018), 23 pages. https://doi.org/10.1145/3161415

[61] Peicheng Yang, Lei Xie, Chuyu Wang, and Sanglu Lu. 2019. IMU-Kinect: A Motion Sensor-Based Gait Monitoring System for Intelligent Healthcare. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) *(UbiComp/ISWC '19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 350–353. https://doi.org/10.1145/3341162.3343766

A SELECTED FEATURES

A.1 Features Used For Risk-prediction Task

[mean_feet_dist, std_left_x, std_imbalance, std_nose_y, max_feet_dist, max_right_x, min_left_x, min_imbalance, range_left_x, range_imbalance, range_nose_x, height _reduction]

A.2 Features Used For Severity-assessment Task:

['mean_feet_dist', 'height_reduction', 'max_feet_dist', 'max_right_x', 'std_imbalance', 'min_left_x', 'range_imbalance', 'range_nose_x', 'mean_nose_y', 'min_nose_x', 'entropy_nose_y', 'std_tilt', 'std_left_x', 'range_feet_dist', 'range_left_x']